

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,300

Open access books available

130,000

International authors and editors

155M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Demographic Analysis and the Decomposition of Social Change

Parfait Eloundou-Enyegue, Sarah Giroux and Michel Tenikue

Abstract

Social science has made great strides over the last half-century, with some of the most significant gains made in micro-level studies. However, analysts interested in broad societal change will not be satisfied with this micro-level detail alone. They will find the detail useful, but they still need to convert the micro-level relations into macro-level outcomes. Decomposition methods rooted in demography can help in those situations. This chapter discusses how these decomposition methods can build on other methods traditionally used in the social sciences. It specifies the kind of problems that are well suited for decomposition analysis, and it briefly reviews three basic types of decomposition approaches (demographic, regression, and mathematical). We illustrate, using mortality data as an example, and conclude with some suggestions for how this method might more broadly advance macrosocial research.

Keywords: decomposition, demographic methods, social change, secondary data, sustainable development

1. Introduction

Social science has made great strides over the last half-century, with some of the most significant gains made in micro-level studies. With advances in computing and communication technology, researchers can now collect, share, and process statistical data on millions of households and individuals. The combination of increased computing power and the expanded availability, size, and complexity of social science datasets has resulted in a “data gold rush,” with the gold digging powerfully aided by remarkable advances in novel statistical methods [1]. With these methods, researchers can now explore in great detail the mix of factors shaping individual behaviors.

These methods can offer powerful insights into individual behavior. Yet, scientists and planners are primarily interested in societal rather than individual outcomes. For example, rather than how individual characteristics –such as education level– shape an individual’s likelihood of smoking, we might also want to see the big picture: how would an educational expansion at the national level shape changes in smoking rates? While the former, microlevel question can be robustly addressed with existing micro-regressions, the latter cannot. This is unfortunate when the interest is in national-level outcomes.

Answering such societal questions is tricky on two fronts. First, as Robinson [2] argued 70 years ago, is the risk of ecological fallacy when one draws macro-level conclusions from micro-level analyses, or vice-versa. Thus, a researcher may find

highly educated individuals to be less likely to smoke, but it does not follow that highly educated *countries* will necessarily have lower smoking rates. Second, the same regression methods cannot robustly apply to countries as units of analysis. Even if we had adequate data for every country in the world (an optimistic scenario), we would still have a relatively small sample of about 200 cases. Such a small sample cannot support robust statistical analyses, especially if one begins to consider many covariates [3, 4]. Despite being widely used in development economics in the 1990s, cross-country regressions have fallen out of favor. Yet, the broad questions about societal transformations remain.

Researchers thus need alternative approaches to link micro-data and macro-issues, building on the detail and robustness of microlevel statistics as they aggregate them to inform macro-level questions. Decomposition methods, a broad set of tools that emerged from the field of demography, present a useful set of options [5–11]. Researchers interested in studying social transformations (spontaneous or induced changes in the structure or performance of a large community or state) can leverage various decomposition approaches to examine issues where the unit of analysis is not the individual but the community.

These methods can be used when the study outcome meets three critical criteria; namely, it is 1) quantifiable, 2) aggregate, and 3) the result of a gradual change. A *quantifiable* outcome is one that can be captured as an absolute number, a percentage, a ratio, or an average. This is the case for outcomes such as rates of smoking or marriage, or average incomes. Second, the phenomenon studied must reflect an *aggregation of individual behaviors*, rather than being an intrinsic/indivisible feature of an entire society. Thus, an outcome such as a country's migration laws (an intrinsic feature with no micro-level correlate) would be excluded. However, one could very well study migration rates among different groups since this is an aggregated measure of individual behaviors. Last, the outcome must be a phenomenon that *changes gradually over time*- such as changes in the ratio of male to female wages. These methods are not suited to sudden changes or rare events (i.e., deaths from Hurricane Katrina) but rather to processes that unfold gradually.

1.1 Limitations and advantages of decomposition methods

Decomposition methods are accounting tools. Thus, while excellent at describing the processes, sectors, or groups driving change, they say little about causation. Thus, in examining changes in smoking rates, decomposition methods could reveal the extent to which different education groups (i.e., no schooling, primary, secondary, higher) contributed to declines in smoking rates. Still, they cannot say *why* education caused smoking rates to change. This is a sizable shortcoming, as causal understanding remains the holy grail for scholars and policymakers seeking to develop better-informed theory and policy. However, understanding the proximate drivers of change is still a useful guide in policy development.

Additionally, decomposition methods have four key advantages over other forms of data analysis: they are 1) easy to interpret; 2) transparent; 3) compatible with other research methods, and 4) efficient. The findings from decomposition methods are easily applied and interpreted because they use basic analyses that do not rely on complex statistics or software. For instance, one standard output from decomposition is to show the percentage of social change accounted for by a given process or group. Such an output (a percentage) is easily digestible compared to outputs from regression analyses such as beta coefficients or odds ratios. The method is transparent because the results of a decomposition analysis are easy to replicate, and the accuracy of the results is easy to check: the sum of all groups' contributions is 100%. The efficiency of decomposition methods reflects its

flexibility and the tradeoff between data requirement, analytical technicity, and the type of findings. Its basic forms can be modified and support more complex combinations to suit the researcher's individual needs. Also, decomposition methods are not very data intensive. In the field of global development, for instance, one can conduct insightful decomposition analyses by relying on tabulations and data sources that are widely available online.

As the examples in this contribution will show, the method and results are quite transparent. First, the input data and sources can be easily checked online by other scholars. Second, unlike multivariate statistics, where results can vary heavily depending on the model specification and individual coding details, decomposition results do not depend on the vagaries of individual modeling choices. Finally, the output from complex regression analysis often seems to be spewed from an impenetrable black box, and it must be taken at face value, with the reader often unable to detect an odd result from, say, a programming mistake. Such is not the case with decomposition: the decomposition findings are presented in a way that allows the reader to immediately assess the internal coherence, credibility, and accuracy of the results.

A third strength is that the decomposition approach can be leveraged and combined with other methods of analysis. In our previous work, we combined a micro-level regression examining the associations between the number of siblings a child has and their educational outcomes, with a decomposition analysis [12]. By combining these two approaches, one can aggregate the robust micro-level findings and answer the macro-level question of how fertility transitions (a country level phenomenon) impact educational attainment at the national level. This integration of methods bypassed the standard limitations of cross-country regression, while also drawing upon the robust micro-level regression findings. The decomposition method is also compatible with many other methods, including qualitative analysis. By quantifying the key behavioral and compositional changes and stratifying across groups and processes, decomposition methods can serve to direct the qualitative research. Essentially, this series of methods does not replace or compete with other methods but, rather, it complements them and expands our toolkit in innovative ways.

Last but not least, decomposition methods efficiently leverage existing information. While some parts of the world are indeed experiencing a “data gold rush,” researchers and policymakers in many parts of the global South often operate with a paucity of data [1, 13]. Despite recommendations by the International Monetary Fund's (IMF) General Data Dissemination System (GDDS) that countries conduct censuses every ten years, 66 countries currently fail to meet the standard [14]. Many researchers in these settings turn to publicly available nationally representative living standards (e.g., the Living Standards Measurement Surveys) or health surveys (e.g., the Demographic Health Surveys or Multiple Indicator Cluster Surveys). These sources can generate robust micro-level data, but they do not occur at regular intervals and are spotty or unavailable for numerous countries. For instance, from 2002 to 2011, 57 countries had either zero or a single poverty estimate [15]. Thus, decomposition methods are an ideal tool that allows scholars to leverage limited data in creative ways.

2. Decomposition basic types

Decomposition is not a single method but a set of related methods. It is used across different fields, but its variants are insufficiently integrated. Most researchers know only the variants that directly apply to their issues of interest, such as life expectancy [8, 16], job discrimination [6, 17], and poverty or inequality

[18, 19]. Yet all these variants can fit into the general taxonomy presented here. Their commonality is in using accounting-based approaches to describe patterns of change. However, they vary in the functional relationship between independent and dependent variables, as shown in **Table 1**.

For each of these relationships, decomposition analysis allows us to examine how a change in the dependent variable is driven by changes in each independent variable (reflecting a group or a process).

2.1 Demographic decomposition

In a demographic decomposition, the main accounting question is about the contribution of ‘composition’ or group-specific size versus ‘group-specific behavior.’ Thus, a researcher might be interested in documenting how much the change in the national rate of smoking between 2000 and 2020 was driven by changes in the distribution of the national population across different education levels

Relationship between X & Y	Example	Nature of the dependent variable	Description	Formula
Demographic	The total smoking rate in a country (Y) is a function of the educational composition of the population (w_j) and the average smoking behavior within each education group (y_j).	Nominal or ordinal (e.g., country region, age group, ethnicity, marital status, educational levels).	The macro-level outcome (Y) is a weighted (by demographic weight, w_j) average of prevailing values in the various subpopulations of the country (y_j)	$Y = \sum (y_j, w_j)$
Statistical	An individual’s expected earnings (Y) as a function of the model intercept (α), the “payoff of each additional year of education (β_t) and the average level of schooling (X_t).	Quantitative (e.g., a person’s years of education, the number of siblings, or income in dollars).	A linear regression relationship between the dependent and independent variables (Y and X, respectively)	$Y_t = \alpha_t + \beta_t X_t$
Mathematical	GDP per capita (Y), which is a function of a country’s GDP (G) and its total population (P).	A quantitative variable (i.e dollars spent per pupil)	The dependent and independent variables are linked by a simple mathematical relationship, which typically involves a quotient, sum, product, or log.	$Y = G/P$

¹Notations used in this table and text are as follows:
X indicates the independent variable.
Y indicates dependent variable for the entire country.
w refers to the demographic weight of an individual subgroup.
 y_j = value of the dependent variable for group j,
 x_j = value of the independent variable for group j.
t = time;
 Δ = indicates the historical change.
 α = intercept or baseline value.
 β = the marginal change in Y associated with a one-unit change in X.
e = error term.

Table 1.
Three basic types of relationships in Decomposition¹.

(compositional effect) versus changes in the smoking rates of each education group (behavioral effect).

Formally, the national rate of smoking rate (Y) is expressed as a weighted average (by w_j) of smoking rates in subpopulations groups defined by educational categories (y_j).

$$Y_t = \sum w_{jt} * y_{jt} \tag{1}$$

In this formula, a national change in the smoking rate can be broken down into two components:

$$\Delta Y = \sum \bar{y}_j * \Delta w_j + \sum \bar{w}_j * \Delta y_j$$

Compositional effect

Behavioral effect

(2)

Formula (2) thus allows the analyst to apportion change into two conceptually compelling components- the compositional ($\sum \bar{y}_j * \Delta w_j$) and behavioral ($\sum \bar{w}_j * \Delta y_j$) effect. The compositional effect is change that is driven by changes in the relative size of each subgroup. The compositional changes are relatively mechanical, as they simply reflect a change in how much each subpopulation is weighted when calculating the national average. The second component of change reflects changes in the behaviors of each subgroup. While the decomposition does not highlight the reasons for change, it does allow the researcher to understand how much of a total change is driven by a meaningful shift in group-specific behavior.

2.2 Regression decomposition

Simple regression analysis tends to model an outcome (Y) as a function of a baseline outcome or intercept (α), a regression coefficient (β), and the value of a predictor variable (X). For instance, one may predict the performance of a well-trained athlete (Y) by knowing the basic performance of a ‘couch-potato’ who has never set foot in a gym (α) and the payoff from each hour of being at the gym (β) multiplied by the number of hours one spends at the gym (X). In that case, if you randomly picked two people in the world, say one from Senegal and the other from Turkey, the difference in their athletic performance (ΔY) can entirely be explained by three factors including:

- the difference in the basic performance of a Senegalese couch-potato vs. a Turkish couch-potato ($\Delta\alpha$)
- the difference in the payoff of gym work in Senegal vs. Turkey ($\Delta\beta$)
- the difference in the number of gym hours for our Senegalese vs. Turkish person (ΔX)

Similarly, a researcher might use a regression decomposition to document how much of the variation in men and women earnings was driven either by the differences in education levels or by the differences in returns to education, suggesting that the gap is driven by discrimination processes. In other words, the researcher would like to know the contribution of (1) the male-female difference in the average level of education, (2) the male-female difference in return to education to

male–female wage differences, and (3) the differences in baseline wages between men and women. These three possibilities are explored below. The formal analysis consists of writing the earning equations for males (m) and females (f) and then taking the difference between these two equations.

$$Y_m = \alpha_m + \beta_m X_m \quad (3)$$

$$Y_f = \alpha_f + \beta_f X_f \quad (4)$$

The decomposition seeks to explain the difference in wages based on the change in the various parameters of the regression equation. This change is expressed as follows

$$\Delta Y = \Delta\alpha + \bar{\beta}\Delta X + \bar{X}\Delta\beta \quad (5)$$

In Eq. (5), the term with upper bar represents the average $\bar{x} = \frac{x_m + x_f}{2}$.

Eq. (5) expresses how much of the pay gap is driven by differences in base salaries ($\Delta\alpha$) versus differences in the levels of schooling (ΔX) and differences in return to education ($\Delta\beta$).

Note that the same procedure can apply to both cross-sectional analysis (the difference between two groups in a given year) and longitudinal analysis (the change experienced by one group between years). The approach is the same; only the interpretations differ.

2.3 Mathematical decomposition

In some cases, an outcome of interest is a function of a set of other variables. For instance, a country's GDP per capita (Y) is measured as the size of the economy divided by the total population. Thus, change in GDP (ΔY) between two time periods can be decomposed into two pieces—the amount of change driven by the growth or contraction in the economy (ΔG) and the change in the total population size (ΔP) or precisely change in the inverse of the population ($\Delta(1/P)$).

$Y = G/P$ is differentiated as

$$\Delta Y = (\bar{G} * \Delta(1/P)) + \left(\left(\frac{1}{\bar{P}} \right) * \Delta G \right) \quad (6)$$

This first decomposition is not very informative. However, one can transform the initial equation into a formula that is slightly longer but conceptually richer:

$$Y = \frac{G}{P} = \frac{G}{A} * \frac{A}{P} \text{ where } A \text{ is the working-age population.}$$

With this transformation, G/A represents the economic productivity of the adult population (or π). This is a conceptually important factor in theories of economic growth. The same is true for the new term A/P (or α), which reflects the ratio of the working-age population to the total population, a core variable in the analysis of demographic dividends. Thus, we now have two theoretically interesting variables (π and α) and can decompose national income in terms of these two variables.

$$\Delta Y = (\bar{\pi} * \Delta \alpha) + (\bar{\alpha} * \Delta \pi) \quad (7)$$

Now, any historical change in GDP (ΔY) can be decomposed into the change in productivity (π) versus the change in the share of the working population (α).

3. Demographic decomposition: a deeper dive

While offering readers a review of different decomposition approaches, we wanted to focus in some detail on demographic decomposition. As noted above, demographic decomposition applies to national outcomes (Y) that are an aggregated result of the outcomes of several subpopulations (y_j), each weighted by its relative size (w_j). Formally:

$$Y_t = \sum w_{jt} * y_{jt} \quad (8)$$

For example, the mortality rate of a country (Y_t) is the weighted average of rates (w_{jt}) in different regions or socioeconomic groups. In this formulation, Y , the dependent variable, is quantifiable and the independent variable X is measured nominally and captured by a set of categories (j). For the demographic decomposition to work, the independent variable must meet four critical criteria: *exhaustiveness*, *distribution*, *variability*, and *relevance*.

Exhaustiveness simply means that each member of the population belongs to one, and only one, of the categories within the independent variable. The set of categories must cover the entire population, and the categories have to be mutually exclusive. Distribution refers to the concept that the number of categories cannot be too few (>2) nor too many. If there are too few categories, as in a variable like a dichotomy, the analysis will not be detailed enough to be informative. Yet, with too many categories (i.e., age in single years), the data will end up spread too thinly.

It is also crucial that the independent variable change over time, i.e., the size of the individual categories comprising the independent variable (the w_j) must fluctuate over time. Without such variation, the compositional effect in the decomposition analysis will always remain zero. Because of this, an independent variable like gender is often not ideal, as sex-ratios in a national population rarely change dramatically over time. Similarly, annual income with set threshold cut-offs (i.e., less than \$20,000; \$20,000–\$40,000; \$40,000–\$60,000; more than \$60,000) could be appropriate, but income quartiles would not, since, by definition, there is no change in the size of each quartile over time.

Lastly, a good independent classification variable should be conceptually relevant to the outcome of interest. Thus, for many phenomena, a strong classification variable might be “region of country” if there has been 1) changes in the size of the population within each region and 2) variation in the Y variable by region. This would be especially appropriate if the variation in Y reflects a scenario where programs and policies are designed at the regional level.

3.1 Two patterns of change: a visual representation

Demographic decomposition allows the analyst to document how much of a particular change is driven by 1) changes in the behavior of different social groups vs. 2) changes in the demographic size of these social groups.

Both compositional and behavioral factors can drive a range of important changes in national social, economic, and health outcomes. **Figure 1** below

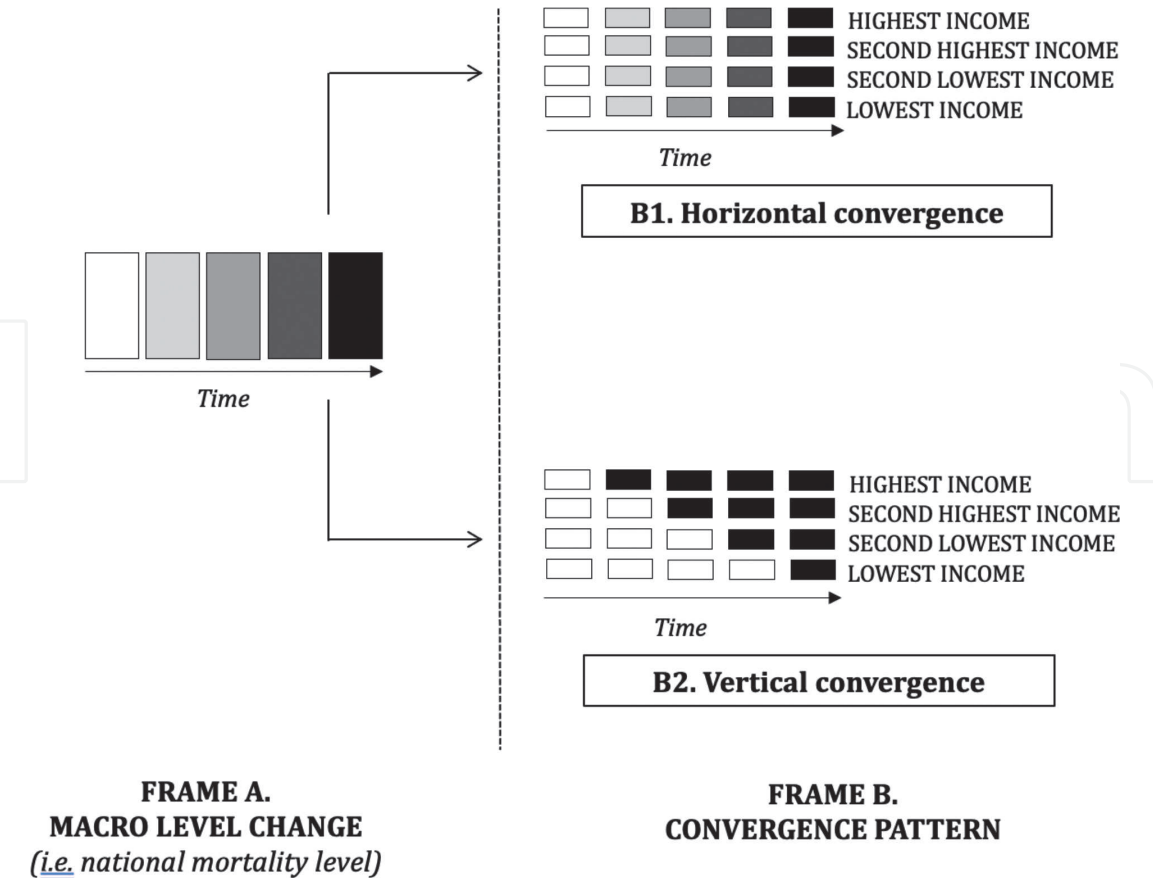


Figure 1.
Patterns of change: Vertical versus horizontal convergence in infant mortality.

illustrates a basic decomposition using the example of infant mortality. This topic is especially relevant for sub-Saharan African policymakers working to achieve Target 3.2 of the Sustainable Development Goals (SDGs), aiming to reduce the levels of child mortality to no more than 25 deaths per 1,000 live births by 2030 [20]. In 2019, Sub-Saharan Africa had the highest neonatal mortality rate at 27 deaths per 1,000 live births [21]. In their first month of life, a child born in sub-Saharan Africa is ten times more likely to die than a child born in a high-income country [21]. However, this fact masks considerable variation within African countries, including severe differences by parental wealth and maternal education [22]. Having a better understanding of the internal variation in these trends can be critical for developing effective policies.

Figure 1 provides a visual description of how the same national-level change can stem from very different subnational patterns. On the left-hand side of **Figure 1** (Frame A: Macro-Level Change) are five squares that represent the trend in child mortality for the country as a whole, with darker colors indicating lower levels of mortality. The chart shows a steady regression at the national level from lower levels of child mortality (the white square on the far left) to higher mortality levels (the black square on the far right). Given these trends, one crucial question is how this evolution occurred, specifically, how various socioeconomic classes contributed to it.

On the right side of the diagram (Frame B) are two possible, and fundamentally opposite, scenarios of how this change can unfold. The first, (B1), highlight a case of horizontal change, where the child mortality rates decline at the same rate for all income groups. Conversely, in Frame B2, we see a vertical change, with the mortality declines beginning with the country's highest income group before gradually spreading to the rest. In year 2, it was only the highest SES group that was

experiencing any decline in child mortality—the other income groups remained unchanged. If we were to ask about the groups that drive the change, we would get different answers from the two scenarios: in the first case, all groups evolved simultaneously, while in the second case, the change occurred first in the higher SES groups. Understanding how this national trend unfolds has key implications for resource allocation and policy targeting.

3.2 Compositional vs. behavioral change

The example above shows how the same national-level change in mortality can emerge from a single group (i.e., entirely driven by mortality declines among the highest SES group) or be something that occurs as the result of a widespread change throughout the population (i.e., all groups experience a decline, but the declines are quantitatively smaller for each group). However, a full decomposition allows the analyst to identify the groups driving the change (referred to here as the compositional effect) and *also* consider how changes in the behavior of each group mattered (referred to here as the behavioral effect).

Figure 2 presents a hypothetical case where a researcher is trying to understand changes in average monthly income. The figure’s left portion shows Time 1, where the average monthly income for all individuals is \$142.50. The classification variable is economic class, ranging from richest to poorest¹, and income is the weighted average of incomes across all of the economic classes making up the national population. On the right-hand side of the figure are two different scenarios, with the average monthly income rising to \$159.20 in both cases. Yet while the two scenarios reflect an identical aggregate change, they are qualitatively quite different.

In scenario 1, the average income of each economic group remains exactly the same between time 1 and time 2. The only factor that changes is the percentage of the population in each economic class. The percentage of the population in the second-highest economic group rose from 15% to 20%, while the poor’s share of the population declined. Thus, the change in scenario one is entirely *compositional*. No individual group became richer, but more individuals moved to higher economic classes. Conversely, in scenario 2, we see the same aggregate gains but no

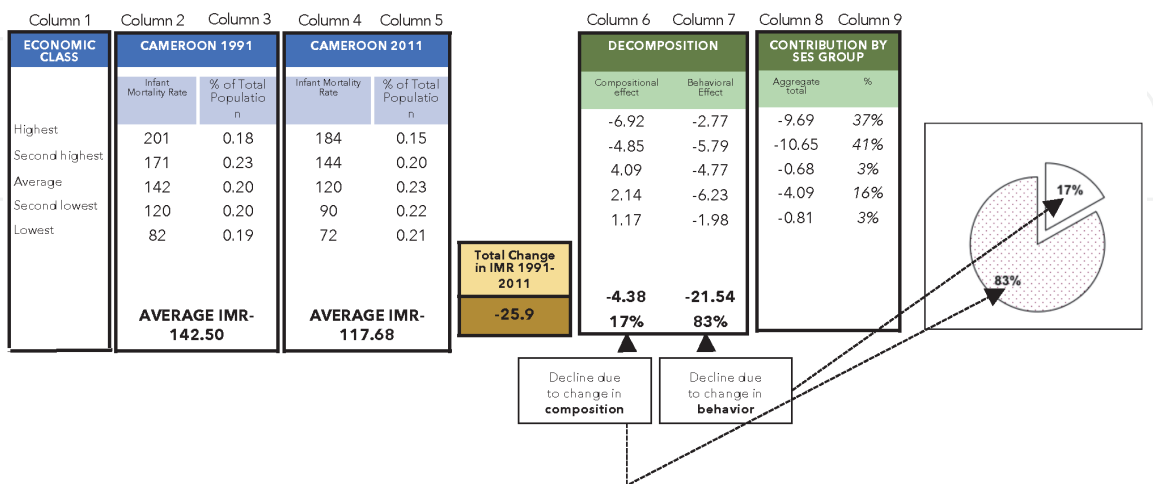


Figure 2.
Decomposing change in child mortality, Cameroon 1991–2011.

¹ For this example, these are not quintiles but based on dollar cutoff values that remain the same between time one and time two.

compositional change. Instead, all of the gains are driven by an increase in the average income among some groups. The wealthiest group became even richer, jumping from \$350 in year 1 to \$400 in year 2.

This example presents an extreme contract, where social change entirely (100%) stems from a change in either composition (Scenario 1) or behavior (Scenario 2). While useful for pedological purposes, it's perhaps unsurprising that reality is usually less extreme. In most cases, change is driven by some combination of the two effects. For example, compositional change might explain 30% of the change, while behavior accounts for the rest. Demographic decomposition allows the analyst to piece apart these different drivers of change (Figure 3).

3.3 Demographic decomposition: the mathematical formulation

In this first case, we focus on a national average; Y will be expressed as a weighted average (by w_j) of the values of individual subpopulations (y_j).

$$Y_t = \sum w_{jt} * y_{jt} \tag{9}$$

In this formula, a national change can be broken down into two components:

$$\Delta Y = \sum \bar{y}_j * \Delta w_j + \sum \bar{w}_j * \Delta y_j \tag{10}$$

As noted above, decomposition methods apportion change into two key components. The compositional effect captures the amount of change driven by changes in each population subgroup's relative size. As seen in the last example, national income may go up simply because the number of individuals in the highest economic class increases, thus increasing the group's demographic weight. Conversely, the behavioral effect captures change that was driven by an actual change in the group's behavior. In our last example, this was reflected by actual gains in income among the lowest economic groups.

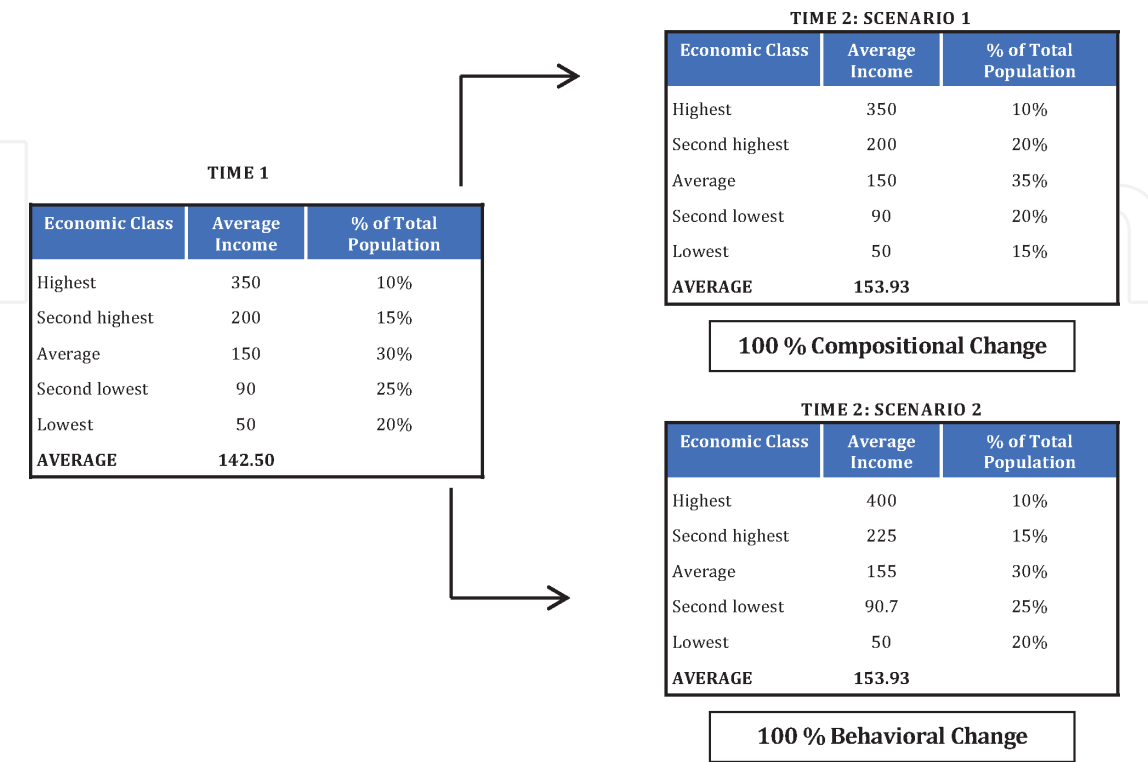


Figure 3. Compositional vs. behavioral change: An extreme case.

In the case of child mortality, declines can similarly be driven by compositional and behavioral effects. Mortality declines could result from an economic contraction whereby the relative size of the lower economic groups, already characterized by higher mortality levels, expands (i.e., their population weight, w , increases). This would reflect a compositionally driven change. Conversely, change can also be driven by changes in the behavior of groups. If one subgroup's mortality levels increase, all else equal, the national mortality will increase. Thus, mortality reversals can also be driven by behavioral change. The researcher's task is to identify how much of the change is driven by each component.

3.4 Key steps

Regardless of the topic, a researcher using demographic decomposition needs to attend to four key tasks, as follows:

1. *Problem definition.* First, the analyst must identify the dependent and independent variables, ensuring that the issues of exhaustiveness, distribution, variability, and relevance (as described above) are suitably addressed. The researcher must then identify the time period they are going to examine. This choice may be constrained by data limitations (i.e., ensuring that the measures are standard across Time 1 and Time 2) but should ideally reflect a conceptually relevant time period. In the example below, we investigate child mortality (dependent variable), with socioeconomic status as the independent (classification) variable. We compare outcomes between 1991 and 2001, as this was a time of severe economic contraction.

2. *National averages:* To conduct the decomposition, we first need to calculate the national averages for the first and last year. We apply formula #9, wherein the national average (Y_t) is a function of the weight of each subgroup at time t (w_{jt}) and the group-specific mean for the dependent variable at time t (y_{jt}). For our example, the group size (w_{jt}) is calculated from a basic frequency analysis for the socioeconomic status variable. The group-specific mortality rates (y_{jt}) are derived from a means comparison, where mortality rates are the dependent variable, and economic class is the independent variable. This information can be calculated from micro data sets using a standard statistical package, but can also come from aggregated reports and studies or online data tools (i.e., the Demographic Health Survey Statcompiler; World Bank Development Indicators; Afrobarometer Online Analysis Tool), making the analysis even more straightforward.

3. *Decomposing the Change.* Once the national averages are calculated for Time 1 and Time 2, the analyst can then use Formula 10 to decompose the change. As the calculations are simply repeated for each group, the researcher can either use Excel software or create a mini-program in a statistical software package to generate the full output. **Figure 2**, below, summarizes the basic data for the calculations.

For our example, the analyst is trying to understand the extent to which changes in the composition of the population (here, changes in the relative size of each economic class) versus behavioral changes of economic groups (here, changes in the mean mortality rate of each economic group) drove the declining rates of infant mortality in Cameroon.

Column 1 highlights the social class categories used (Highest, Second Highest, Average, Second Lowest, Lowest). Columns 2–5 are where the analyst must insert their data (gathered elsewhere) on group size (w_{jt}) and behavior (y_{jt}). The spreadsheet is built to then calculate the national average for each year, here 143.6 in 1991 to 117.7 in 2011. The resulting difference reflects a decline in infant mortality of 25.9.

The last step is then to explain this 25.9 unit decline, using formula 10. As reflected in Column 6, the researcher can see that 17% of the decline was driven by compositional change. This is unsurprising, as the class composition of the population changed slightly during this period. The proportion of children living in low-income families declined (from 18% to 15%) while there was a slight rise in the proportion of children living in rich families (19 to 21%). As poorer families exhibit higher infant mortality rates, mortality rates will mechanically go down if their representation in the population shrinks. Column 7 displays the amount of change due to behavioral changes, which was the larger contributor to change, at 83%. A glance at Columns 2 and 5 (showing mortality rates by economic class in 1991 and 2011) confirms that mortality rates did indeed decline across all social classes during this period, including the poorest.

4. *Presentation of findings.* **Figure 2** is easily presentable to most scientific audiences, requiring little explanation. Moreover, the researcher can not only highlight the leading processes- i.e., how much of the change was driven by compositional (Column 6) vs. behavioral (Column 7) forces- but can also explore leading groups. The percentages in column 9 are simply the result of adding the behavioral and compositional contributions of each socioeconomic class and dividing by the total change. The second-lowest economic group's contribution was the largest, with the total change equal to the sum of its composition effects (-4.85) and behavior (-5.79) for a total of -10.65 , or 41% of the total change. The lowest economic group made the next largest contribution to the mortality decline (37%), followed by the second highest (16%), average (3%), and highest (3%) groups.

As is evident from **Figure 2**, the sum of all contributions in a decomposition reflects the total change- thus, the values in columns 6 and 7 sum to -25.9 , the total decline in mortality. As a percentage, the sum of all contributions must equal 100%. One important note is that some contributions can be negative (less than 0%) or greater than 100%. Conceptually, a negative value reflects a contribution that worked in the opposite direction of the observed change. In our case, a negative percentage would reflect a change in either the composition or behavior of a group that worked to increase mortality. Percentages larger than 100% signal that the overall change would have been even greater, if it was not hampered by the effects of opposite influences from other groups.

While relatively easy to digest, decomposition findings are easy to present to non-technical audiences as a graph rather than a table. Pie charts, such as the one on the left side of **Figure 2**, can cleanly show how much of the total change was driven by compositional vs. behavioral factors. Similarly, stacked histograms can efficiently summarize the results and identify the dominant social groups driving change.

4. Policy implications

Even if decomposition is not a causal method, it can guide policy. For instance, if a planner observes that much of recent national change is derived from

compositional changes, then s/he gains insights into future changes. In a country where education levels are rising, having a positive compositional effect (having more educated people leads to progress), then one can anticipate further gains if national levels of education continue to grow. Likewise, if child mortality is declining mostly via a compositional effect (fewer high-parity births), then one expects child mortality to continue to decline if fertility does. In some cases, the planner may have reason to expect the composition changes to continue largely on their own, as is the case of mortality, educational, fertility transitions. In some instances, s/he must proactively induce compositional change through policy.

If mortality is driven by a behavioral effect, the appropriate response is to target either the leading group or the lagging group. One would target the leading groups if one is not worried about inequalities and one does not expect further growth in this top group to be curtailed by a ceiling effect. One has further justification for investing in this leading group if one expects the example set by this vanguard group to trickle down and promote change among the following groups. Some development theorists might argue that in the early stages of development, it can make sense to build up some pioneers who would set the pace and pull the rest of the population [23]. On the other hand, one may favor the lagging group if one assumes that leading already has a momentum on their own and will continue to progress even if unaided. One may further favor the lagging groups out of concern for inequality.

This application of decomposition to policy requires a nuanced theoretical understanding of processes of change and the diffusion of innovations in the general population. In diffusion theory, change may accelerate among the lagging groups after reaching a critical mass in preceding groups [24, 25]. In other ways, change proceeds in a domino pattern and, to the extent that decomposition analysis helps identify the next group in that domino line, it can speed the process.

5. Conclusion

Students of social change need robust methods to inform policy reliably. With the significant advances in data and computing technology achieved over the last forty years, they are in a better position to study micro-level processes, including the causes of individual behavior. However, this microlevel expertise is not sufficient to account for social change, where the focus is on aggregate (not individual) outcomes. Applying the existing micro-methods methods to understand aggregate social problems amounts to “barking up the wrong tree” or—to stay with this canine metaphor- “letting the methodological tail wag the dog.” Decomposition methods can help address this micro-macro conundrum by making it possible to aggregate evidence from smaller units to understand the big picture.

Given that the big picture, rather than individual-level detail, is the focus of most socioeconomic planning, decomposition methods are quite relevant to policy. The methods are ideal for studying many of the social transformations underway across the globe. In particular, they can inform the study of critical components in the United Nations’ Sustainable Development Goals (SDG), including poverty, health, inequality, and schooling. Many countries are working hard to achieve these goals face with severely limited data and resources. A fuller understanding of the drivers of socioeconomic change and the unevenness of change in these rapidly changing and diverse societies can allow policymakers to target policies more effectively. To this end, decomposition methods can help.

Acknowledgements

The authors would like to acknowledge support from the Hewlett Foundation and the Minerva Program in the US Department of Defense. We also thank colleagues and students in our Demographic Methods course at Cornell (DSoc 4080/PAM 6060) for their valuable comments.

Author details


Parfait Eloundou-Enyegue^{1*}, Sarah Giroux¹ and Michel Tenikue²

¹ Department of Global Development, Cornell University, Ithaca, NY, USA

² Luxembourg Institute for Socioeconomic Research (LISER), Esch-sur-Alzette, Luxembourg

*Address all correspondence to: pme7@cornell.edu

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Felt M. Social media and the social sciences: How researchers employ Big Data analytics. *Big Data Soc* 2016; 3: 2053951716645828.
- [2] Robinson WS. Ecological correlations and the behavior of individuals. *Stud Hum Ecol Row Peterson Evanst Ill* 1961; 115–120.
- [3] Levine R, Zervos SJ. What we Have Learned about Policy and Growth from Cross-Country Regressions? *Am Econ Rev* 1993; 83: 426–430.
- [4] Poll M. *Breaking Up The Relationship: Dichotomous Effects of Positive and Negative Growth on the Income of the Poor*. Centre for the Study of African Economies, University of Oxford, 2017.
- [5] Kitagawa EM. Components of a difference between two rates. *J Am Stat Assoc* 1955; 50: 1168–1194.
- [6] Oaxaca R. Male-female wage differentials in urban labor markets. *Int Econ Rev* 1973; 693–709.
- [7] Gupta PD. *Standardization and decomposition of rates: A user's manual*. US Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1993.
- [8] Vaupel JW, Romo VC. Decomposing change in life expectancy: A bouquet of formulas in honor of Nathan Keyfitz's 90th birthday. *Demography* 2003; 40: 201–216.
- [9] Gupta PD. A general method of decomposing a difference between two rates into several components. *Demography* 1978; 15: 99–112.
- [10] Gupta PD. Decomposition of the difference between two rates and its consistency when more than two populations are involved. *Math Popul Stud* 1991; 3: 105–125.
- [11] Oaxaca RL, Ransom MR. Identification in detailed wage decompositions. *Rev Econ Stat* 1999; 81: 154–157.
- [12] Eloundou-Enyegue PM, Giroux SC. Fertility transitions and schooling: From micro-to macro-level associations. *Demography* 2012; 49: 1407–1432.
- [13] Ye Y, Wamukoya M, Ezech A, et al. Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Saharan Africa? *BMC Public Health* 2012; 12: 741.
- [14] Chandy L, Zhang C. *Stuffing data gaps with dollars: What will it cost to close the data deficit in poor countries?* Brookings Institution, <https://www.brookings.edu/wp-content/uploads/2016/06/0827-Chandy-Data-financing-gapfinalv3.pdf> (August 2015, accessed 17 January 2021).
- [15] Serajuddin U, Uematsu H, Wieser C, et al. *Data Deprivation: Another Deprivation to End*. The World Bank. Epub ahead of print 28 April 2015. DOI: 10.1596/1813-9450-7252.
- [16] Cui Q, Canudas-Romo V, Booth H. The Mechanism Underlying Change in the Sex Gap in Life Expectancy at Birth: An Extended Decomposition. *Demography* 2019; 56: 2307–2321.
- [17] Afridi F, Dinkelman T, Mahajan K. Why are fewer married women joining the work force in rural India? A decomposition analysis over two decades. *J Popul Econ* 2018; 31: 783–818.
- [18] Shorrocks A, Wan G. Spatial decomposition of inequality. *J Econ Geogr* 2005; 5: 59–81.
- [19] Filauro S, Parolin Z. Unequal unions? A comparative decomposition of income inequality in the European

Union and United States. *J Eur Soc Policy* 2019; 29: 545–563.

[20] United Nations Population Division. *Population Facts*. New York, NY: United Nations Department of Economic and Social Affairs, https://www.un.org/en/development/desa/population/publications/pdf/popfacts/PopFacts_2017-6.pdf (December 2017, accessed 26 January 2021).

[21] World Health Organization. Newborns: improving survival and well-being, <https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality> (2020, accessed 26 January 2021).

[22] Anyamele OD, Ukawuilulu JO, Akanegbu BN. The Role of Wealth and Mother's Education in Infant and Child Mortality in 26 Sub-Saharan African Countries: Evidence from Pooled Demographic and Health Survey (DHS) Data 2003–2011 and African Development Indicators (ADI), 2012. *Soc Indic Res* 2017; 130: 1125–1146.

[23] Deaton A. *The great escape: health, wealth, and the origins of inequality*. Princeton University Press, 2013.

[24] Rogers EM, Shoemaker FF. *Communication of Innovations; A Cross-Cultural Approach*.

[25] Kapoor KK, Dwivedi YK, Williams MD. Rogers' innovation adoption attributes: A systematic review and synthesis of existing research. *Inf Syst Manag* 2014; 31: 74–91.